Received: 28 July 2021 | Revised: 2 June 2022 | Accepted: 6 June 2022

DOI: 10.1111/desc.13296

Developmental Science WILEY

RESEARCH ARTICLE

Learning how to learn from social feedback: The origins of early vocal development

Steven L. Elmlinger | **Jennifer A. Schwade** | **Laura Vollmer** | **Michael H. Goldstein**

Department of Psychology, Cornell University, Ithaca, New York, USA

Correspondence
Michael H. Goldstein, Department of Psychology, Cornell University, 211 Uris Hall, Ithaca, NY 14853, USA.
Email: michael.goldstein@cornell.edu

Funding information
National Science Foundation, Grant/Award Number: BCS-0844015

Abstract

Infants' prelinguistic vocalizations reliably organize vocal turn-taking with social partners, creating opportunities for learning to produce the sound patterns of the ambient language. This social feedback loop supporting early vocal learning is well-documented, but its developmental origins have yet to be addressed. When do infants learn that their non-cry vocalizations influence others? To test developmental changes in infant vocal learning, we assessed the vocalizations of 2- and 5-month-old infants in a still-face interaction with an unfamiliar adult. During the still-face, infants who have learned the social efficacy of vocalizing increase their babbling rate. In addition, to assess the expectations for social responsiveness that infants build from their everyday experience, we recorded caregiver responsiveness to their infants' vocalizations during unstructured play. During the still-face, only 5-month-old infants showed an increase in vocalizing (a vocal extinction burst) indicating that they had learned to expect adult responses to their vocalizations. Caregiver responsiveness predicted the magnitude of the vocal extinction burst for 5-month-olds. Because 5-month-olds show a vocal extinction burst with unfamiliar adults, they must have generalized the social efficacy of their vocalizations beyond their familiar caregiver. Caregiver responsiveness to infant vocalizations during unstructured play was similar for 2- and 5-month-olds. Infants thus learn the social efficacy of their vocalizations between 2 and 5 months of age. During this time, infants build associations between their own non-cry sounds and the reactions of adults, which allows learning of the instrumental value of vocalizing.

KEYWORDS
communicative development, parent-infant interaction, social learning, vocal learning

1 | INTRODUCTION

The prelinguistic vocalizations of infants are potent signals which drive early vocal turn-taking with social partners. Babbling elicits bouts of spontaneous vocal turn-taking with caregivers, providing infants with opportunities for communicative learning from social responses to their vocalizations (Albert et al., 2018; Gratier et al., 2015; Jaffe et al., 2001). Human infants have a long period of vocal immaturity, during which vocal development seems particularly open to social input (e.g., Goldstein & Schwade, 2010; Ramírez-Esparza et al., 2017). By 9 months, infants produce more speech-like vocalizations as a result of caregiver contingent feedback on their vocalizations (Goldstein & Schwade, 2008; Goldstein et al., 2003). In addition, caregivers of 9-month-old infants simplify the content of their speech that is contingent on babbling, creating a favorable environment for vocal learning (Elmlinger et al., 2019, 2019b). Thus, vocal development is a socially situated process, characterized by a feedback loop in which babbling both influences and is influenced by social interaction (Goldstein & Schwade, 2010; Warlaumont et al., 2014).

How does the feedback loop itself develop? An infant's first year of life involves extensive changes in vocal production (Oller, 2000). Within the first 3 months, infants' early non-cry vocalizations (squeals, growls and vowel-like sounds) are produced in a functionally flexible manner such that sounds are not coupled to specific states or objects (Jhang & Oller, 2017; Oller et al., 2013). Although these early vocalizations are not yet speech-like, their functional flexibility affords infants the exploration of both the form and function of their sounds (Oller et al., 2019). In light of these findings, most research on infant vocal development focuses on acoustic changes, termed *production learning* (Goldstein & Schwade, 2008; Janik & Slater, 2000; McLeod & Crowe, 2018; Vihman, 2017). Production learning is sensitive to the social environment because infants reorganize their vocal repertoires as they are influenced by their exposure to lexical forms of the ambient language (Ha et al., 2021). Infants' vocal repertoires are sensitive to social input, and infant vocalizations reliably elicit changes in the immediate social environment (Elmlinger et al., 2021; Faust et al., 2020; Goldstein & Schwade, 2010).

How do infants learn that their own vocalizations elicit learning opportunities from social contexts? The answer to that question requires studying the effects of social responsiveness on vocal production in younger infants, before they are producing canonical syllables or first words. Within the first 6 months, infants exhibit *contextual learning* in which infants learn that their babbling elicits contingent, prompt and positive social responses from caregivers (Goldstein & Schwade, 2010; Janik & Slater, 2000). Thus, a critical task of prelinguistic vocal development is to learn the affordances of social contexts that facilitate advances in vocal development. In our view, the social-vocal feedback loop rests on the developing awareness that social partners' responses contain useful information about the communicative force of infants' early babbling. The present study investigates this foundational form of contextual vocal learning.

Evidence for such learning comes from infants' reaction when social responses are withheld. Five- and 6-month-old infants temporarily increase their rate of vocalizing when adult social responses cease (Bourvis et al., 2018; Franklin et al., 2013; Goldstein et al., 2009). Results from human and non-human animals demonstrate that once associations between a behavior and its result are learned, removing the contingency between behavior and consequence causes distinct changes in behavior (McConnel & Miller, 2014). After the contingency is removed, the rate of the specific behavior which elicited the contingency temporarily increases and then immediately decreases. For example, if children find a previously functional marble dispenser suddenly inoperable, they react by pressing the previously functional button with greater intensity (Holton, 1961). This pattern of increase followed by decrease is called an "extinction burst" because an extinction of an expected consequence elicits a burst of behavior (Amsel, 1958; Dunsmoor et al., 2015; McConnel & Miller, 2014). Infants' temporary increase in vocal activity when social feedback is extinguished–a vocal extinction burst (VEB)–likely reflects that infants have learned the social efficacy of their vocalizations (Figure 1). The VEB is an outcome of instrumental vocal learning, as infants have specifically associated their own vocal actions with social response outcomes. When the learned behavior no longer results in the event, the learner first increases the production of the behavior and then immediately decreases (Mackintosh, 1975). While 1- to 3-month-old infants vocalize during a cessation of adult responding, whether their vocalizations follow a burst pattern is unknown (Bigelow & Power, 2016).

While many studies have focused on prelinguistic vocal learning and phonological development, much less attention has been paid to the even earlier process of contextual vocal learning. In the present study, we tested 2- and 5-month-olds in a still-face paradigm to assess developmental changes in the sensitivity of early babbling to perturbations of social responsiveness. The still-face paradigm has been widely used in studies of social and emotional development (e.g., Adamson &

RESEARCH HIGHLIGHTS

- Infants' knowledge of the social efficacy of their vocalizations gradually emerges over the first 5 months of life.
- Caregiver responsiveness to infants' vocalizations is linked to infants' developing expectation that their vocalizations have social effects on novel social partners.
- We provide evidence of a candidate mechanism of infants' learning how to learn early pragmatic and social precursors to engagement in mature conversation.
- We propose a new theoretical framework connecting vocal and social development based on infants' construction of social expectations.

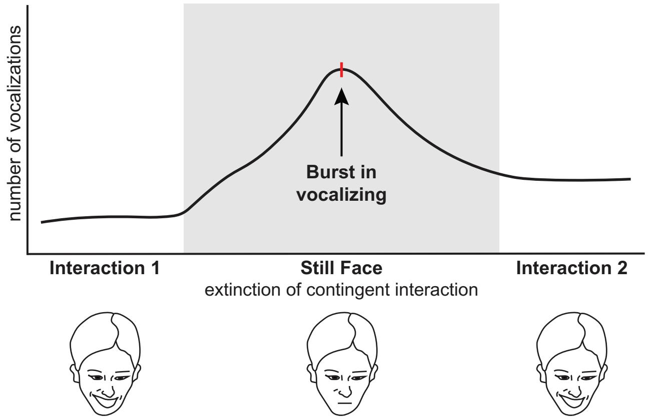

FIGURE 1 Depiction of the vocal extinction burst (VEB), in which extinction (i.e., cessation) of social responsiveness during the still face period results in a burst of infant vocalizing. The red vertical dash indicates peak vocalizing, after which point vocalizing decreases. The VEB is calculated by subtracting the number of infants' vocalizations per minute during Interaction 1 from their vocalizations per minute during Still Face. Under the *social expectancy hypothesis*, the VEB is driven by infants' expectations that their vocalizations elicit changes in caregiver behavior based on their prior experiences in naturalistic interaction

Frick, 2003; Mesman et al., 2009). The paradigm consists of a social interaction between an infant and an adult. After a brief period of naturalistic interaction, the caregiver becomes silent and maintains a neutral expression while looking at the infant (Bornstein et al., 2004; Tronick et al., 1979). After 2 min, the caregiver resumes typical interaction (Bornstein et al., 2004). During all three episodes, infant behavior is observed to determine whether the infant has detected a change in social responsiveness during the still face. Infants are typically monitored for changes in affect and self-regulation (e.g., self-touching and averting gaze from the caregiver). At the onset of the still-face episode, infants between 1 and 4 months often orient toward their mother and smile. This initial increase in orienting and smiling is interpreted as an attempt by the infant to reengage the social partner (Tronick et al., 1979). From about 1.5 months of age, infants respond to the continued lack of social interaction during the still-face episode with gaze aversion and decreased smiling (Adamson & Frick, 2003; Striano & Bertin, 2005). Six-month-old infants typically show logarithmic decreases in gaze to parent and smiling across the still face episode (Ekas et al., 2013).

We hypothesize that the sensitivity of vocal production to changes in the social context is grounded in the development of infants' predictions about the social effects of their vocalizations (Faust et al., 2020; Pickering & Garrod, 2013). Each infant brings to the still face paradigm a months-long developmental history of social interaction that has afforded opportunities for learning and the building of expectations for future social engagement. Grounding early vocal learning in the imperfect contingencies of social interaction may provide an explanation for the large individual differences in the VEB that were observed at 5 months (Goldstein et al., 2009) and 6 months (Bourvis et al., 2018; Franklin et al., 2013). A review of the still face paradigm illustrated relations between maternal sensitivity during the first episode of naturalistic interaction and infants' visual gaze and emotional reactions during the following still-face episode (Mesman et al., 2009). For example, maternal sensitivity to 6-month-old infants during the first interaction period is correlated with more positive infant behaviors during the still face period (Tronick et al., 1982). Such correlations may be a product of the immediately preceding maternal interaction or may result from the infant-caregiver dyads' history of interaction (Bigelow et al., 2018; Mesman et al., 2009). While social interaction history is thus acknowledged as a potent force in socio-emotional development, it has received much less attention in the domain of early vocal learning.

In our view, the vocal extinction burst (VEB) is built on infants' social expectations that they have formed with their caregiver through everyday social interaction. Learning simple associations between one's own babbling and contingent social responses allow the infant to gradually form a generalized prediction that their vocalizations elicit responses from their social environment. We refer to this idea as the *social expectancy hypothesis* (Figure 1).

As the VEB is a key indicator of contextual vocal learning, what is the developmental trajectory of the VEB? We hypothesize that the VEB should emerge gradually because the social contingencies that afford prediction are imperfect. During in-home naturalistic interaction, approximately 30% of vocalizations obtain verbal responses from caregivers (Fagan & Doveikis, 2017). In non-vocal domains, for example, head movement to activate a mobile (Watson, 1967) or arm movement to trigger music (Sullivan & Lewis, 2003), learning from such imperfect contingencies develops slowly. Infants show evidence of learning from imperfect contingencies between their own behavior and external events around 4–5 months of age (Sullivan & Lewis, 2003). If early vocal learning from social feedback is based on similar general mechanisms of learning (Tarabulsy et al., 1996), then infants may not be able to learn the social efficacy of their non-cry vocalizations until 4–5 months of age. Further evidence suggesting a later onset of infant vocal learning from social contingencies comes from the development of social smiling. Although the onset of social smiling occurs at around 2 months of age, infants increase their social smiling from 2 to 5 months (Messinger, 2005; Messinger et al., 2010). The control of social smiling may develop in contingent social interactions similar to those that support vocal learning (Ruvolo et al., 2015). The gradual development of vocal learning in social contexts may constrain infants' experience of receiving responses from people beyond their immediate caregivers.

When in development do infants gain the ability to generalize the social efficacy of their vocal behaviors beyond their immediate caregivers? In the present study, experimenters participated in the still face procedure instead of the infants' caregivers. Evidence suggests by 2 months of age that infants readily differentiate their mother's voice, face and responsiveness from that of a stranger's (Bigelow & Rochat, 2006; DeCasper & Fifer, 1980; Field et al., 1984; Orena & Werker, 2021). Thus, infants have very early data from exposure to caregiver speech, faces, and behavior that informs discriminating familiar from unfamiliar adults. Once infants form expectations about their caregiver's responsiveness, a subsequent challenge of communicative development is the generalization of social expectations to the behavior of unfamiliar social partners, thus utilizing past experience in the service of future communication (Tenenbaum & Griffiths, 2001). While 5-month-old infants clearly come to expect that their vocalizations engage caregivers, the extent to which younger infants generalize their social expectations to unfamiliar adults is presently unknown.

In sum, data from multiple domains suggests that infants' early social experience facilitates the development of social expectations that guide vocal learning. From birth, infants are capable of forming associations between their own behavior and subsequent environmental events (DeCasper & Spence, 1986; Vouloumanos & Werker, 2007). In vocal learning, early associations between babbling and social responses provide data informing the developmental construction of predictions about the social efficacy of vocalizing. Previous research with children has established the extinction burst as a measure of prediction strength, in which the magnitude of a child's burst response is predicted by the strength of their prediction for a given outcome to occur (Holton, 1961). The more history children had with a marble dispenser functioning properly, the more forceful their subsequent button-presses after it ceased to function (Holton, 1961). Similarly, infants' predictions organize their vocal behavior when social contingencies are removed, resulting in attempts to reestablish the predicted outcome by increasing the vocalization rate. This attempt is the vocal extinction burst.

In contrast, a purely associative account of vocal learning would not predict an extinction burst in vocalizing. After removal of a social contingency, an associative model would predict only subsequent decreases in vocal behavior (Rescorla & Wagner, 1972). However, generalized expectations can emerge from action-consequence contingency learning (see review in Maier & Seligman, 2016). Animals form expectations about future events based on prior contingencies. Recent computational work has provided evidence that school-aged children can track their expectations about the *controllability* of environments (Raab et al., 2022). That is, the extent to which one's own actions predict state changes in the environment above and beyond previous states reflects the controllability of an environment. Incorporating the efficacy of one's own actions into their predictions about the world is an efficient way to guide future learning (Dorfman & Gershman, 2019; Ligneul, 2021). In this view, infants' burst in vocal activity could be a result of attempts to reduce the difference between expected controllability of social responses and observed lack of controllability during the still-face period. Thus, we believe that action-consequence contingency learning mechanisms account for the developmental emergence of expectations for the social efficacy of babbling.

To better understand the social and developmental origins of contextual vocal learning, in the present study we used a still-face paradigm to elicit a vocal extinction burst in 2- and 5-month-olds. We tested whether the level of maternal responsiveness to prelinguistic vocalizations during naturalistic play predicted individual differences in the VEB. We hypothesized that 5-month-olds would show a VEB but that 2-month-olds would not. If 2-month-olds have not yet learned the social efficacy of their babbling, then their rate of vocalizing should not change across the still-face periods.

While our hypothesized changes in the VEB are the result of learning the social efficacy of babbling, we also assessed smiling as a measure of infants' noticing that social responding had ceased during the still-face. By 2 months of age, infants show sensitivity to reductions in social contingency by decreasing smiling (Nadel et al., 1999). Thus, we predicted that 2-month-olds would respond to the still-face episode with decreased smiling. However, under the *social expectancy hypothesis*, we predicted that they would not yet show evidence of instrumental vocal learning because their limited experience with social contingency would not permit learning relations between their own vocal behavior and social responses. This result would be consistent with findings of gradual development in social smiling. We hypothesized that the level of maternal responsiveness to prelinguistic vocalizations would be positively related to the magnitude of the VEB, as infants who receive more contingent maternal responses to their vocalizations are more likely to predict that their vocalizations will have instrumental social value during the still face interaction.

In summary, we hypothesize that if 2- and 5-month-olds have learned the social efficacy of their non-cry vocalizations, they will exhibit a VEB, and if 2- and 5-month-olds detected a lack of positive affect in the still-face period, they will decrease their smiling behavior. If the *social expectancy hypothesis* holds, we predict that caregiver responsiveness specifically to infant vocalizations in free play will predict the VEB.

2 | METHODS

2.1 | Participants

Forty mother-infant dyads participated in the study. We tested 20 infants aged 2 months (mean age 2;9, range 1;28–2;27) and 20 infants aged 5 months (mean age 5;10, range 4;19–5;17). All infants were healthy and full-term. The final samples were approximately balanced for infant gender (2 months: eight males, 12 females; five months: 10 males, 10 females). Parents received an infant t-shirt or bib for their participation. An additional 25 infants were tested but excluded from the final sample for the following reasons: crying or excessive fussiness (2-month-olds: $n = 7$, 5-month-olds: $n = 16$), falling asleep during the still-face interaction (2-month-olds: $n = 1$), and parental interference during the still-face interaction (5-month-olds: $n = 1$). Attrition rates were similar to those from other studies employing the still-face paradigm (e.g., Hsu & Jeng, 2008 observed 20% attrition with 2-month-olds; Yirmiya et al., 2006 observed 42% attrition with 4-month-olds).

2.2 | Apparatus

The study took place in a 3.7 m × 5.5 m room containing toys, picture boards, and an infant seat. The session was recorded using three wall-mounted, remotely controlled video cameras routed to a digital tape deck (Panasonic AGDV2000) via a video mixer (Videonics MXProDV). Infant vocalizations were recorded by a wireless microphone (Telex FLM-22) and transmitter (Telex USR-100) that was carried in denim overalls worn by the infants. The microphone was contained in a pocket in the front of the overalls. The wire and transmitter were concealed in the lining of the overalls and did not impede infants' movement. Infant vocalizations were routed to the left stereo channel of the video tape recorder via an audio mixer (Mackie 1604VLZ). Mothers' and experimenters' speech were recorded by a wireless microphone (Telex FLM-22) and transmitter (Telex FMR1000) worn in a pouch at the adults' waist. The microphone was attached to the adult's collar. During the still-face interaction, the experimenter wore a pair of wireless headphones (Radio Shack model 33–1196) so that an observer in the control room could cue the onset of each episode. The observer's instructions were recorded and routed to the experimenter's wireless headphones and to the right stereo channel of the video tape recorder.

2.3 | Procedure

The procedure included an unstructured play period followed by a still-face interaction. At the beginning of the play period, mothers were instructed to play as they would at home for 10 min. The still-face interaction always followed the play interaction because infants often become fussy during still-face interactions (Delgado et al., 2002). If the still-face interaction occurred first, carryover effects from the still face might change mothers' typical patterns of responsiveness to infants, as responsiveness may change when infants are fussy or have recently been fussy. After the play period, infants participated in a 4-min still-face interaction with an experimenter. An experimenter was used rather than the caregiver for the still-face interaction to provide a higher level of control and consistency over the adult behavior presented to the infants.

Our procedure follows that of a previous study, in which 5-month-old infants demonstrated a significant vocal extinction burst during a still face episode while interacting with an unfamiliar adult (e.g., Goldstein et al., 2009). The still-face interaction contained three episodes: a naturalistic interaction episode (Interaction 1: 1 min), a still face episode (2 min), and a second naturalistic interaction episode (Interaction 2: 1 min; Bornstein et al., 2004; Goldstein et al., 2009). During each interaction episode, the experimenter engaged the infant in a face-to-face interaction. She spoke to the infant in an animated way but did not touch the infant or engage the infant with any toys. During the still-face episode, the experimenter maintained a neutral expression while looking silently at the infant. During the still-face interaction, the infant was seated in an infant seat on the floor. The experimenter knelt on the floor in front of the infant so that her face was at infants' eye level. Caregivers sat in a chair behind and out of view of infants while completing a demographic questionnaire. The still-face interaction was videotaped using a split screen, with one camera focused on the experimenter and a second camera focused on the infant. The still-face interaction was conducted by one of four female experimenters throughout the study. Experimenters who conducted the still-face interaction with infants had not previously interacted with the infant. The use of experimenters in the still-face interaction allowed us to assess the extent to which infants generalized the social effects of their communicative behaviors to novel social partners.

2.4 | Coding and analysis

We counted the frequency of infant vocalizations during the play interaction and during each episode of the still-face interaction. Each syllable (any vocalization containing a vowel) was counted as a separate vocalization (e.g., Goldstein et al., 2003). For example, [dada] would be counted as two vocalizations. Vocalizations could also be separated by a breath (e.g., [a] < breath > [a] would be counted as two vocalizations). Fusses, cries, and vegetative sounds such as coughs were excluded (Oller, 2000). Smiles were coded frame-by-frame and were defined as occurring whenever an infant contracted their zygomatic muscle to raise either one or both lip corners (Jones et al., 1991). To assess changes in infant vocalizations across the still-face interaction, we calculated a difference score for vocalization frequency per minute during the still-face episode minus vocalization frequency during the first naturalistic interaction episode.

During the play interaction, we coded mothers' responses to their infants' vocalizations. To provide a baseline level of maternal responsiveness, we also coded maternal responses to instances of infants' dyadic non-vocal behavior. We categorized infants' behavior as dyadic non-vocal if the infants smiled at their mothers, turned to look at their mother's face or touched their mother (Gros-Louis et al., 2006). A maternal response was scored if a mother changed her behavior (e.g., by touching, speaking to, or smiling at her infant) within 5 s of the offset of an infant behavior. Our previous work showed that nonverbal responses to babbling facilitate vocal learning at similar rates as verbal responses (Goldstein & Schwade, 2008; Goldstein et al., 2003), thus we report all contingent responding to infant behavior. Maternal responses that could not be detected by the infant were not coded (e.g., if an infant sat with her back to the mother and the mother smiled). A 5 s threshold for detecting responses was used in accord with previous studies of communicative development in 2- and 5-month-olds and older infants (e.g., Donnellan et al., 2020; Gilkerson et al., 2017). The proportion of maternal responses to infants' vocalizations and non-vocal behavior were then calculated separately as the number of infant behaviors in a category that elicited a maternal response divided by the total number of infant behaviors in that category.

Infant vocalizations and maternal behavior were scored using custom-designed software that allowed for frame-accurate annotation of events in digital video and audio (EventCoder; Goldstein & Brodsky, 2006). The play interaction and still-face interaction were coded separately to avoid possible coder bias. Behavior during the play interaction was initially scored by one of two coders. Twenty percent of the play interactions were independently re-scored by another coder to assess reliability. Reliability was calculated with interclass correlations ($r = 0.91$ for maternal behaviors during the play session, $r = 0.99$ for infant vocalizations during the play session). Infant vocalizations and smiles during the still-face interaction were also initially scored by one of two coders. Twenty percent of the still-face interactions were independently re-scored by another coder to assess reliability. For behaviors during the still-face interactions, the interclass correlation was excellent for both infant vocalizations ($r = 0.96$) and infant smiles ($r = 0.97$).

3 | RESULTS

3.1 | Vocalizations during the still-face interaction

We used a linear mixed effects model to test for differences in infants' vocalizations per minute as a function of still-face episode (using the *lme4* package in R; Bates et al., 2015). In this model, the fixed effect was the still-face episode (Interaction 1, Still Face, Interaction 2) per age group (2- and 5-month-olds), and the random effect was individual subject. We compared models using likelihood ratio tests to determine whether inclusion of a fixed effect and interaction made significant contributions to the model. There were no significant main effects of episode (χ^2 (2) = 2.87, $p = 0.2379$) or age (χ^2 (1) = 1.10, $p = 0.2923$), but the episode by age interaction trended significant (χ^2 (2) = 4.72, $p = 0.0940$). As 2-month-olds likely have greater volubility than 5-month-old infants, the overall effect of episode may not be readily comparable across age groups (Iyer et al., 2016). As the inclusion of the episode by age interaction into the model improved fit, and due to the known differences in 2- and 5-month-old infants' volubility, we further assessed vocalizations within each age group.

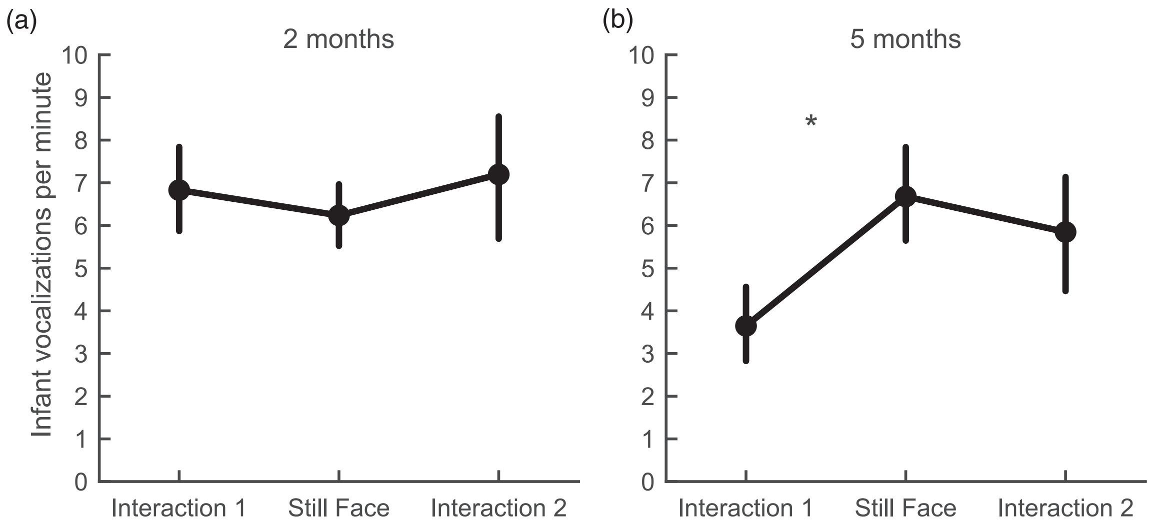

FIGURE 2 Mean vocalizations per minute during each episode of the still-face interaction, by age: (a) 2 months, (b) 5 months. Error bars represent $\pm$ 1 SE. * $p < 0.05$

TABLE 1 Results of generalized linear mixed model (LMM) to test whether the number of vocalizations per minute and smiles per minute changed with still-face episode

Difference between episodes	Estimate (SE)	t-value	Effect size	adjusted p-value
2 months vocalizations per minute ($n = 20$ infants)				
Interaction 1 – Still Face	0.59 (1.19)	0.50	0.15	0.8724
Interaction 2 – Still Face	0.96 (1.19)	0.81	0.25	0.7008
Interaction 1 – Interaction 2	−0.37 (1.19)	−0.31	−0.09	0.9489
5 months vocalizations per minute ($n = 20$ infants)				
Interaction 1 – Still Face	−3.03 (1.19)	−2.54	−0.80	0.0397 *
Interaction 2 – Still Face	−0.83 (1.19)	−0.69	−0.22	0.7666
Interaction 1 – Interaction 2	−2.19 (1.19)	−1.84	−0.58	0.1696
2 months smiles per minute ($n = 20$ infants)				
Interaction 1 – Still Face	3.34 (0.76)	4.42	1.39	0.0002 ***
Interaction 2 – Still Face	1.21 (0.76)	1.61	0.50	0.2550
Interaction 1 – Interaction 2	2.12 (0.76)	2.81	0.88	0.0206 *
5 months smiles per minute ($n = 20$ infants)				
Interaction 1 – Still Face	2.04 (0.77)	2.64	0.83	0.0308 *
Interaction 2 – Still Face	2.78 (0.77)	3.61	1.14	0.0024 **
Interaction 1 – Interaction 2	−0.74 (0.77)	−0.96	−0.30	0.6007

SE are model estimated standard errors.
The first column indicates which two periods of the still face interaction were being compared in the model.

To decompose the effect of episode in 2- and 5-month-old infants, separate models were constructed per age group, each using model parameters as above. Two-month-old infants' vocalizations per minute did not show a significant effect of episode (χ^2 (2) = 0.69, $p = 0.7072$; Figure 2a; Table 1). During the Interaction 1 episode, 2-month-olds vocalized marginally more than 5-month-olds (3.18 $\pm$ 1.64 (*SE*); $t = 1.94$, $p_{adj} = 0.0554$). No other between-age comparisons were significantly different (*ts* < 0.82, *ps* > 0.4124).

It was possible that 2-month-old infants (who tend to have higher volubility than older infants; e.g., Iyer et al., 2016) did not change their vocal activity over episodes because their vocalizing was already at ceiling during Interaction 1. However, half of the 2-month-old infants showed an increase in vocalizations per minute from Interaction 1 to the Still Face episode ($n = 10$ of 20, Wilcoxon signed-rank test, $V = 124$, $p = 0.4980$), indicating that these infants were capable of increasing their volubility in the Still Face period. The increase shown by the 2-month-olds, however, was not distributed as an extinction burst (see *Changes in infant vocalizations* below and Figure 4a). In contrast, 5-month-old infants' vocalizations per minute showed a significant effect of episode (χ^2 (2) = 6.66, $p = 0.0357$). Post-hoc tests correcting

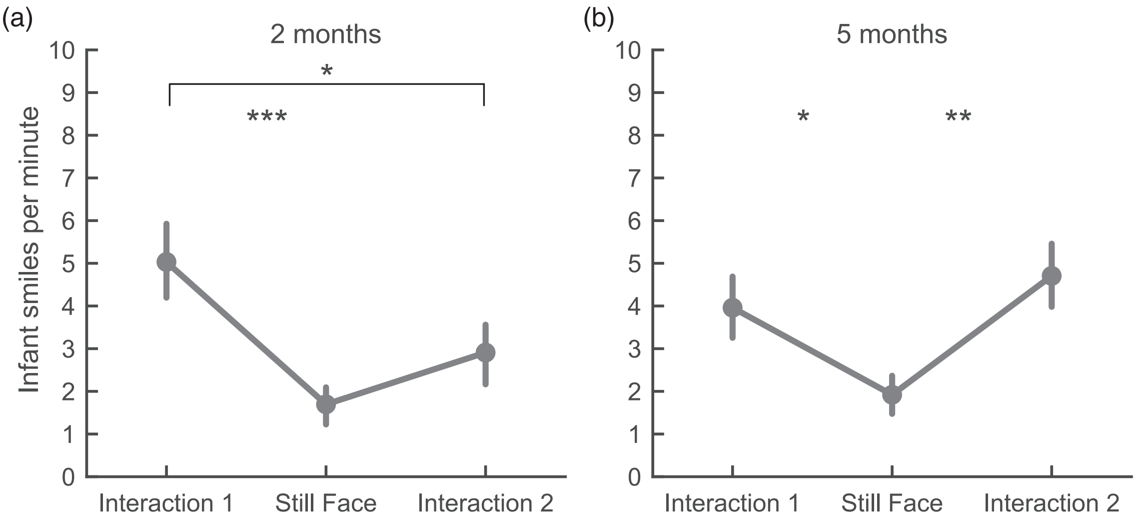

FIGURE 3 Mean smile per minute during each episode of the still-face interaction, by age: (a) 2 months, (b) 5 months. Error bars represent ± 1 SE. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

for multiple comparisons were derived using the *emmeans* package in R (Lenth et al., 2018). A Tukey-adjusted p-value method for multiple comparisons revealed that 5-month-olds vocalized more during Still Face than during Interaction 1 (Figure 2b; Table 1). A significant number of 5-month-olds showed an increase in vocalizations per minute from Interaction 1 to the Still Face episode (12 increased, one did not change, and seven decreased, Wilcoxon signed-rank test, $V = 41$, $p = 0.0313$).

3.2 | Smiles during the still-face interaction

To test for differences in infants' smiles per minute as a function of still-face episode (Interaction 1, Still Face, Interaction 2) per age group (2- and 5-month-olds) we used a mixed model as in the above vocalization analysis. The dependent variable in this model was the number of infant smiles per minute. As shown in Figure 3, there was a significant main effect of episode (χ^2 (2) = 22.37, $p < 0.0001$), no main effect of age (χ^2 (1) = 0.19, $p = 0.66$), but a significant trial by age interaction effect (χ^2 (2) = 7.14, $p = 0.0280$).

To decompose the interaction, separate models were constructed per age group, each using model parameters as above. Two-month-old infants' smiles per minute showed a significant effect of episode (χ^2 (2) = 16.92, $p = 0.0002$). A Tukey-adjusted p-value method for multiple comparisons revealed that 2-month-olds smiled less during Still Face than during Interaction 1 and also smiled less during Interaction 2 than in Interaction 1 (Figure 3a; Table 1). Five-month-old infants' smiles per minute showed a significant effect of episode (χ^2 (2) = 12.56, $p = 0.0018$). Tukey-adjusted p-values revealed that 5-month-olds smiled less during Still Face than during Interaction 1 and also smiled more during Interaction 2 than in Still Face (Figure 3b; Table 1). No other between age comparisons were significantly different (*ts* < 1.11, *ps* > 0.2698).

3.3 | Changes in infant vocalizations and smiles during the still-face interaction

To assess changes in vocalizing during the still-face episode, we divided the 2-min episode into eight 15-s periods, following the procedure of Goldstein et al. (2009) and Franklin et al. (2013). For each age group, we compared the mean number of vocalizations per 15-s period to the last 15 s of the first interaction episode using single-sample t-tests with Bonferroni corrections (corrected $\alpha = 0.05/8 = 0.00625$). Within the 2-min Still Face episode, the number of 2-month-old infants' vocalizations did not differ from baseline in any 15 s period, *ps* = 0.288 – 0.999 (Figure 4a). In contrast, 5-month-old infants' vocalizations increased to the midpoint of the Still Face episode (Figure 4b). The number of vocalizations in the midpoint time bin (from 120 to 135 s) was significantly above baseline, $t(19) = 3.11$, Bonferroni-corrected $p = 0.046$.

To assess changes in infant smiling during the still-face episode, we divided the episodes into 15-s periods as in the above vocalization analysis. We compared smile frequency per 15-s period to the last 15 s of the first interaction episode with Bonferroni corrected t-tests, following Goldstein et al. (2009). Within the 2-min Still Face, 2-month-olds' smiling reached a minimum at 135 s (Figure 5a). 2-month-olds' mean number of smiles was significantly below baseline from 75 s through the end of Still Face, *ts* (19) < −3.53, Bonferroni-corrected *ps* < 0.05. Within the 2-min Still Face, 5-month-olds infants' smiling reached a minimum at 120 s (Figure 5b). 5-month-olds' mean number of smiles was significantly below baseline from 105 to 135 s, $t(19) < -4.70$, Bonferroni-corrected *ps* < 0.01.

3.4 | Infant behavior and caregiver responses during free play

To assess infants' behavior during free play with caregivers, we calculated the number of infant vocalizations per minute, and the number

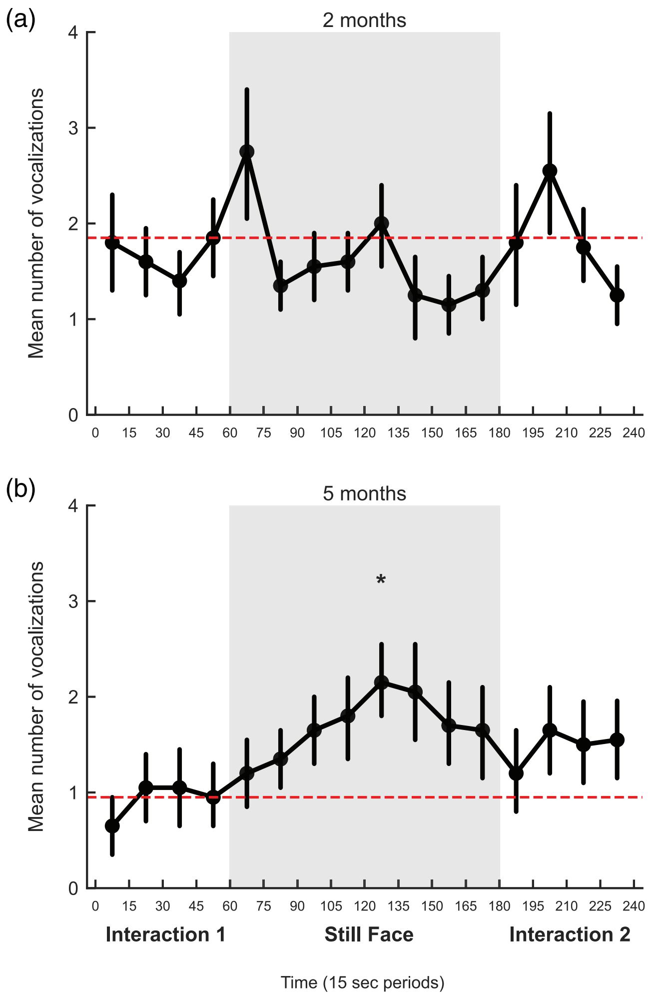

FIGURE 4 Mean vocalization frequency during the still face episode, in each 15 s period. Error bars represent ± 1 SE. (a) 2-month-old infants (b) 5-month-old infants. The grey shaded area from 60 to 180 s represents the duration of the still face episode. The red dashed line indicates baseline, defined as mean vocalization frequency during the last 15 s of the first naturalistic interaction. Mean vocalization frequency in each 15 s period was compared to baseline with Bonferroni-corrected t-tests. * Bonferroni-corrected $p < 0.05$

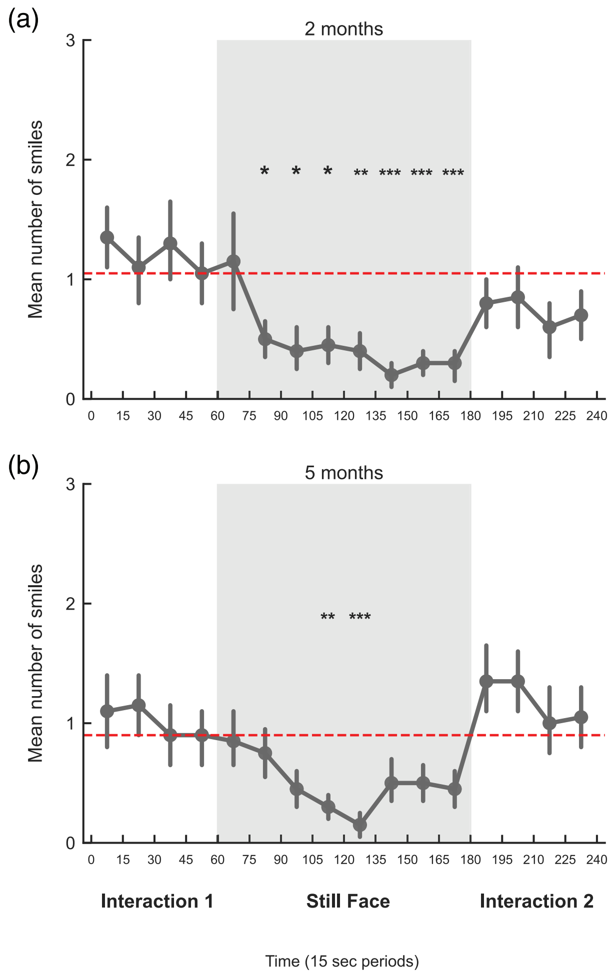

FIGURE 5 Mean smile frequency during the still face episode, in each 15 s period. Error bars represent ± 1 SE. (a) 2-month-old infants (b) 5-month-old infants. The grey shaded area from 60 to 180 s represents the duration of the still face episode. The red dashed line indicates baseline, defined as mean smile frequency during the last 15 s of the first naturalistic interaction. Mean vocalization frequency in each 15 s period was compared to baseline with Bonferroni-corrected t-tests. *** Bonferroni-corrected $p < 0.001$, ** Bonferroni-corrected $p < 0.01$, * Bonferroni-corrected $p < 0.05$

of infants' dyadic non-vocal behaviors per minute across the duration of their free play session (M_{dur} = 10 min, 7.93 s, SD_{dur} = 10.66 s). Two-month-olds vocalized 5.76 (SD = 3.62) times per minute on average, while 5-month-olds vocalized 3.15 (SD = 2.38) times per minute on average. These vocal rates are comparable to previous descriptions of infants' volubility during free play with caregivers (Iyer et al., 2016). Two-month-olds engaged in dyadic non-vocal behaviors 0.74 (SD = 0.86) times per minute while 5-month-olds did so 0.96 (SD = 0.84) times per minute.

To gauge caregivers' natural responses to their infants' behaviors during free play, we derived the proportion of infants' behaviors (vocalizations and dyadic non-vocal behaviors) which elicited a caregiver response. 74% (SD = 12%) of 2-month-old infants' vocalizations elicited a caregiver response, while 67% (SD = 18%) of 5-month-old infants' vocalizations did. These response elicitation rates are consistent with previous reports of caregiver responses to young infants' vocalizations during unstructured play in laboratory settings (Albert et al., 2018; Gros-Louis et al., 2006). 74% (SD = 35%) of 2-month-old infants' dyadic non-vocal behaviors elicited a caregiver response, while 60% (SD = 34%) of 5-month-olds elicited a response.

To test whether infants inhibited their vocalizations during the still-face interaction, we compared the rate of vocalizations (calculated as number of vocalizations per minute) in the free play session to the rate of vocalizations in the first two periods of the still face interaction. Older infants vocalized more during the Still Face episode than during the free play session ($M_{Still\ Face\ -\ free\ play}$ = 3.42, SD = 2.38), $t(19) = 2.89$, $p = 0.0093$. There was no significant difference in rate of vocalizing between 5-month-old infants during their Interaction 1 episode and their play session, $p = 0.7830$. For the 2-month-old infants, there was

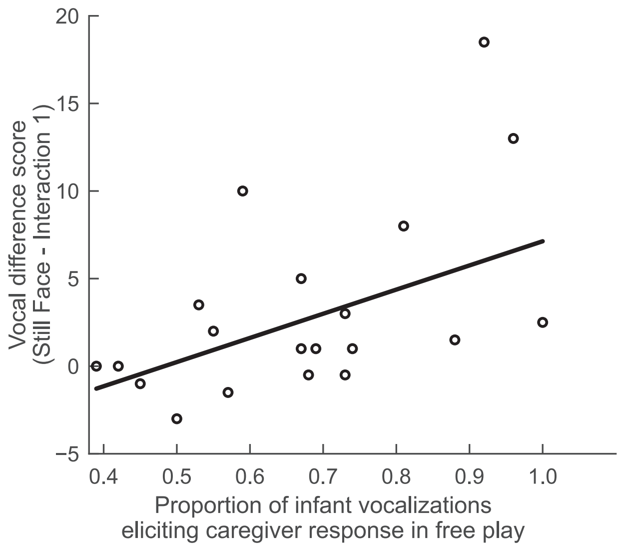

FIGURE 6 Relationship between the magnitude of the 5-month-old infants' vocal extinction burst (VEB) in the still-face procedure and infants' vocalizations during free play which elicited a caregiver response ($r(18) = 0.55$, $p = 0.011$)

no significant difference between rate of vocalizing in the play session and either still-face episode, *ps* > 0.5090.

3.5 | Testing the social expectancy hypothesis by comparing behavior between free play and still-face

The social expectancy hypothesis predicts that the magnitude of individual infants' vocal extinction burst (VEB) may be predicted by the extent to which infants' vocalizations reliably elicit caregiver responses during free play (Figure 1). To test this hypothesis, we investigated the relation between the magnitude of the VEB during the still-face and caregiver responsiveness to infant vocalizations during free play. We correlated the still-face difference score (calculated by subtracting the number of infants' vocalizations per minute during Interaction 1 from their vocalizations per minute during Still Face) with the proportion of infant vocalizations which elicited caregivers' response during the free play interaction. For 2-month-old infants, the magnitude of the VEB was not significantly correlated with caregiver responsiveness to vocalizations during free play, $r\,(18) = -0.02$, $p = 0.913$. In addition, the VEB was not significantly correlated with caregiver response to dyadic nonvocal behaviors, $r\,(18) = -0.08$, $p = 0.707$.

In contrast, for 5-month-old infants, the magnitude of the VEB was positively correlated with the proportion of infant vocalizations which elicited a caregiver response during free play, $r\,(18) = 0.55$, $p = 0.011$ (Figure 6). We found no significant correlations for either 2- or 5-month-old infants between vocalizations per minute in any still face period and caregiver responsiveness to vocalizations in free play (all *ps* > 0.07). Five-month-olds' VEB was not correlated with the proportion of infants' dyadic non-vocal behaviors which elicited a caregiver response, $r\,(18) = -0.18$, $p = 0.43$. These results are consistent with the hypothesis that infants' expectation for social responses following their own vocalizations predicts the magnitude of their VEB (Figure 1).

Infants may have exhibited vocal bursts by chance during the still-face period. To control for this possibility, we assessed the relation between volubility during free play and the magnitude of the VEB. As before, the VEB was calculated by subtracting the number of infants' vocalizations per minute during Interaction 1 from their vocalizations per minute during Still Face. We found no significant relations between the frequency of infants' vocalizations per minute during any of the still-face episodes, including the VEB, and 2- and 5-month-old infants' volubility during unstructured play with caregivers, *rs* $(18) < 0.18$, *ps* > 0.433. Specifically, 5-month-old infants' volubility (vocalizations per minute) during the entire free play period did not correlate with their VEB during still-face, $r(18) = -0.06$, $p = 0.801$. Thus the VEB was unlikely to be due to chance bursts of vocalizing.

4 | DISCUSSION

Between 3 and 5 months of age, infants learn that their prelinguistic vocalizations change the behaviors of social partners. Our findings show that by 5 months, infants come to expect their caregivers to be responsive to their vocalizations, and they generalize this expectation to novel social partners. Five-month-olds significantly increased non-cry vocalizations from the first naturalistic interaction episode to the still-face episode and then trended in a decreasing direction until the second naturalistic interaction episode. Infant vocalizations increased to a peak at the midpoint of the 2-min Still Face episode, followed by a decrease to the end. The initial increases in vocal activity, immediately followed by a decrease is characteristic of a vocal extinction burst (VEB). This result is consistent with previous findings of contextual vocal learning in 5-month-olds (Goldstein et al., 2009) and 6-month-olds (Bourvis et al., 2018; Franklin et al., 2013).

The expectation for caregiver responsiveness in 5-month-olds appears to be learned during everyday social interaction when caregivers promptly and contingently respond to their infants' non-cry vocalizations (Ritwika et al., 2020). Individual differences in 5-month-olds' vocal extinction bursts during the still-face were positively correlated with caregiver responsiveness to infant vocalizations during the free play interaction (Figure 6). Although caregivers also responded frequently to infants' non-vocal dyadic behavior (e.g., infants' smiling), responsiveness to non-vocal behavior did not predict the magnitude of the VEB. Thus, infants' expectations were specific to the effects of their vocalizations. In addition, infants' volubility during free play with caregivers did not predict their VEB.

In contrast to the instrumental vocal learning shown by the 5-month-olds, 2-month-old infants did not change their rate of vocalizing during the Still Face episode. These younger infants do not yet show evidence of instrumental vocal learning. In addition, there was no relation between the 2-month-olds' rate of vocalizing during the Still Face period and caregiver responsiveness to their vocal or non-vocal behavior, even though caregivers of 2- and 5-month-olds showed similar levels of responsiveness to prelinguistic vocalizations.

The reduction in smiling in the still-face interaction indicates that both 2- and 5-month-olds recognized their social partners' unresponsiveness and decreased their positive affect (Northrup et al., 2019). Why did infants not use smiling as a way to reestablish responsiveness? There are at least two possibilities for the lack of burst in infants' smiling during the still-face. One possibility is that the temporal structure of typical interaction consists of temporally overlapping smiling behaviors with fewer opportunities for learning contingencies, as in vocal turn-taking where overlapping interruptions are rare (Bigelow et al., 2018; Levinson, 2016). The temporal structure of reciprocal vocal turn-taking may afford more opportunities for learning the social efficacy of vocalizations than does mutual smiling. Another possibility is that the extinction of a social partner's positive affect overrides any infant propensity for smiling (Messinger, 2002). While infants use vocalizations to express a range of emotional content, infant smiles may be relatively constrained to positive affect contexts (Oller et al., 2013).

The overall pattern of results suggests that the emergence of instrumental vocal learning occurs between 2 and 5 months of age. Our results point to infants' vocal interactions with caregivers as an important source of early individual differences in learning the social efficacy of their sounds. These findings give developmental context to previous work that has shown a VEB in 5-month-olds (Goldstein et al., 2009) and 6-month-olds (Bourvis et al., 2018; Franklin et al., 2013).

Taken together, our results are consistent with the *social expectancy hypothesis*. Instrumental vocal learning – learning that vocal behavior has social effects – emerges over time, with social contingency perception as an important mechanism. The development of infants' instrumental vocal learning is supported by the incremental associative learning of expectations that their vocalizations will have prompt effects on social partners. Through many iterations of everyday vocal turn-taking with caregivers, infants build predictions about social responses to their vocalizations. Thus, a precursor to an advanced pragmatic and social skill – influencing others – is learned through the accumulation of simple social contingencies. Initially, social influences on vocal learning are limited by infant's restricted ability to learn from the imperfect contingencies that characterize parents' reactions to infant vocal behavior. Evidence from non-social paradigms suggests similar constraints and show that learning of imperfect contingencies gradually increases from 2 to 3 months of age (Wentworth & Haith, 1992). Thus, the 2-month-olds in the present study either cannot learn or have not had enough opportunities to learn the social efficacy of their vocalizations, given caregivers who respond with typical contingency levels. A likely precursor to contextual vocal learning is the presence of 2-month-old infants' social bidding during the still-face, defined as visually attending to the mother while also smiling or making non-distress vocalizations (Bigelow & Power, 2016; Nadel et al., 1999). Such social bidding was predicted by maternal responsiveness to smiles and vocalizations during the interaction period prior to the still-face (Bigelow & Power, 2016). In the present study, with changes in vocalization amount from interaction to still-face as the primary measure, instrumental vocal learning was predicted by the responsiveness of the social environment specifically to non-cry vocalizations.

In our view, a primary task of early communicative development is learning the instrumental value of vocalizing. The *social expectancy hypothesis* also illustrates how the early developmental challenge of learning how to learn the social power of one's voice gives rise to the social feedback loop that characterizes more advanced vocal learning over the first year. Participating in the social feedback loop facilitates sociopragmatic skills in three ways. First, imperfectly predictable environments organize infants' attention (Kidd et al., 2012; 2014). The social feedback loop is characterized by imperfectly predictable social responses, which organize infants' attention around moments when adults produce mature speech (Miller & Gros-Louis, 2013; Miller et al., 2009). Second, as infants refine their predictions about the social consequences of vocalizing, their more accurate predictions likely facilitate more accurate acoustic representations of mature caregiver speech. A candidate mechanism of this is the activation of dopaminergic reward circuitry when predictions increase in accuracy (Schultz et al., 1997). Reward signals in cortico-basal ganglia circuits are thought to be an important component of speech production learning (Ackermann et al., 2014; Guenther, 2016; Syal & Finlay, 2011). Recent studies have shown that dopaminergic reward from increasing prediction accuracy can facilitate learning about the sensory features of what is being predicted (Sharpe et al., 2017). Third, infants' production of more mature vocalizations facilitates caregiver responsiveness (Abney et al., 2016; Albert et al., 2018; Goldstein & West, 1999). By responding to more mature infant vocalizations over infants' development, caregivers provide acoustic scaffolding for their infants' vocal development (Goldstein & Schwade, 2010).

When infants learn that they are active agents in a social feedback loop, they have achieved a sophisticated step towards rapid, mature communication, in which online predictions about social responses temporally unfold simultaneously with ongoing vocal production (Corps et al., 2017; Levinson, 2016). Thus, the prelinguistic VEB should have predictive power for later developments in communication and language. Indeed, infants' development of expectations for the social consequences of their vocalizing predicts advances in prelinguistic vocal learning and language in the second year. The magnitude of 5-month-olds' vocal extinction burst is positively related to infants' receptive language development at 13 months (Goldstein et al., 2009). Our findings suggest that caregiver responsiveness to infant vocalizations may facilitate both prelinguistic and early language learning, once infants can perceive the imperfect contingencies that characterize social interaction. As vocal development proceeds, it continues to be influenced by the social environment. Caregiver responsiveness to the babbling of 9-month-olds predicts later language development at 13–15 months (Goldstein & Schwade, 2010; Rollins, 2003; Tamis-LeMonda & Bornstein, 2002; Tamis-LeMonda et al., 2001).

In our view, vocal turn-taking with caregivers serves to gradually construct infants' expectations for rapid social effects of their own vocal activity. In their everyday learning environment, contingencies between infants' own behavior and environmental changes are imperfect and context dependent. In-home recordings of 4- to 14-month-old infants show that caregivers verbally respond to approximately 30% of infants' vocalizations (Fagan & Doveikis, 2017). These imperfect contingencies are slowly learned over the first several months. In 4-month-olds, imperfect contingencies are learned at a lower rate than perfect contingencies between infants' arm movements and music stimuli (Sullivan & Lewis, 2003). In the domain of vocal development, prelinguistic vocalizations reliably organize their caregivers' behavior (e.g., Albert et al., 2018; Elmlinger, Schwade et al., 2019, 2019b; Goldstein & West, 1999; Gros-Louis et al., 2006; Kremin et al., 2021; Northrup et al., 2019; Papoušek & Papoušek, 1989; Warlaumont et al., 2014). The moments after an infant vocalization are marked by distinct affective and linguistic changes in caregivers. For example, caregivers, even in non-Western populations, simplify the syntactic and lexical complexity of their speech in response to immature vocalizations (Elmlinger et al., 2022). Thus, the statistics for infants to form seemingly sophisticated assumptions about the efficacy of their vocalizations are present in their early learning environment.

Though we showed the efficacy of imperfect contingencies on learning, we also showed a positive relationship between caregiver responsiveness and the VEB. Even at higher response rates, caregiver behavior will generally not rise to 100% contingency. From the infant's point of view, even high levels of responsiveness are still imperfect. Infants whose parents respond at higher rates are being given more evidence of the social efficacy of their vocalizations. Such parental input builds infants' expectations for future responses, and the extent to which infants entrench their expectations for the social efficacy of vocalizing is captured by infants' reaction to their expectations being violated.

Not all vocalizations have equal social potency. The directedness of infants' vocalizations predicts the structure of caregivers' immediate response (Albert et al., 2018). Few studies have assessed caregiver response rates to infant vocalizations which are specifically directed to caregivers. A recent longitudinal study revealed that infants' caregiver-directed vocalizations that elicit responses are an important predictor of early word production (Donnellan et al., 2020). Future studies should investigate the relations between caregiver-directed vocalizations, caregiver responsiveness, and infants' instrumental vocal learning.

By 5 months, infants appear to be generalizing their expectations for the social effects of their vocalizations beyond their immediate caregivers. For generalization to be possible, infants must first be able to distinguish between their caregivers and unfamiliar adults – a prerequisite that 2-month-olds typically meet (Bigelow & Rochat, 2006; DeCasper & Fifer, 1980; Field et al., 1984). In our view, infants of this age have not been exposed to enough vocal turn-taking with social partners to form expectations of social responsiveness to their vocalizations. In contrast, 5-month-olds have learned that their vocal activity predicts adult responses (both caregivers and unfamiliar adults; Bourvis et al., 2018; Franklin et al., 2013; Goldstein et al., 2009). Findings of a VEB with unfamiliar adults is evidence that infants have generalized this prediction.

What underlying mechanisms connect the coordinated emergence of social and vocal learning? Over the course of development, infants are intrinsically motivated to reduce their uncertainty about the world (Oudeyer & Smith, 2016). This uncertainty reduction is both rewarding to infants, and influential over what is learned next (Wade & Kidd, 2019). When 10-month-old infants engage in a novel social interaction, they appear to find the contingent responses of a non-humanoid robot rewarding, as they allocate more attention to its behaviors and are more likely to follow its "gaze" than do infants to a non-contingent robot (Movellan & Watson, 1987). The extent to which infants find a learning event rewarding is likely predicted by the learning rate or learning progress accrued from the event (Kaplan & Oudeyer, 2004). When a learning event is too complex or already well-understood, the expectation for potential reward is low because the event offers minimal learning opportunities (Kidd & Hayden, 2015). Evidence regarding the role of reward in speech development comes from fMRI recordings which show that children's individual differences in functional connectivity between speech perception and reward circuitry predicts their social-communicative outcomes (Abrams et al., 2013; Abrams et al., 2016; Abrams et al., 2019). The 2-month-olds in our study are not likely to have encountered enough social response opportunities to make learning progress associating their own vocalizations with changes in adult behavior. Infants' learning progress associating their vocalizations with social responses has been argued to drive further engagement with social stimuli over the course of development (see Masek et al., 2021 for thorough review).

There are advantages to learning from less predictable, imperfect contingencies. Learning from variable, imperfect predictability allows for longer maintenance of behavior in the absence of feedback (Amsel, 1958, 1962; Bacon, 1962; Festinger, 1961; Mackintosh, 1975). Vocal patterns learned under imperfect predictability will remain robust to disruptions of interaction and/or caregiver distractions that might create increased inconsistencies or gaps in their responsiveness. In addition, studies of infants' detection of statistical patterns in phonology and speech input have demonstrated that infant learning and generalization are facilitated by the inclusion of variability in the learned input (Gómez & Maye, 2005; Rost & McMurray, 2009; Vukatana et al., 2015).

In summary, our understanding of early communicative development is enhanced by connecting vocal reactions to the still face to characteristics of the social environment. Our finding that mothers of younger and older infants showed similar levels of responsiveness to vocalizations supports our claim that changes in infant learning occur between 2 and 5 months. We believe that early vocal development is initially constrained by infants' restricted ability to learn from the imperfect contingencies of typical social interaction. As the ability to perceive and learn from imperfect contingencies develops, vocal development becomes a socially situated process, related to maternal responsiveness, that allows instrumental vocal learning and eventually speech usage to occur. Current studies in our laboratory are investigating the relative contributions of caregiver responsiveness and domain-general learning mechanisms (i.e., social and non-social contingency perception) to instrumental vocal learning and speech precursors to determine the forces driving vocal development.

ACKNOWLEDGMENTS

We thank the families who participated in the study. Fern Baldwin, Casey Berson, Rachel Brandstadter, Mi Hae Chung, Melissa Frankel, Emily Laucks, Maryam Sajed, and Alissa Worly assisted in conducting the study and coded the behavioral data. We thank Steve Robertson and Kim Oller for their insightful feedback on earlier versions of the manuscript. Data collection was supported by NSF grant BCS-0844015 to MHG.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ETHICAL STATEMENT

Written informed consent was obtained from the parent before the study began. The research protocol was approved by the Cornell University Committee on Human Subjects (approval #06-10-016).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Michael H. Goldstein https://orcid.org/0000-0001-6672-3752

REFERENCES

Abney, D. H., Warlaumont, A. S., Oller, D. K., Wallot, S., & Kello, C. T. (2016). Multiple coordination patterns in infant and adult vocalizations. *Infancy*, *22*(4), 514–539.

Abrams, D. A., Chen, T., Odriozola, P., Cheng, K. M., Baker, A. E., Padmanabhan, A., Ryali, S., Kochalka, J., Feinstein, C., & Menon, V. (2016). Neural circuits underlying mother's voice perception predict social communication abilities in children. *Proceedings of the National Academy of Sciences*, *113*(22), 6295–6300.

Abrams, D. A., Lynch, C. J., Cheng, K. M., Phillips, J., Supekar, K., Ryali, S., Uddin, L. Q., & Menon, V. (2013). Underconnectivity between voice-selective cortex and reward circuitry in children with autism. *Proceedings of the National Academy of Sciences*, *110*(29), 12060–12065.

Abrams, D. A., Padmanabhan, A., Chen, T., Odriozola, P., Baker, A. E., Kochalka, J., Phillips, J. M., & Menon, V. (2019). Impaired voice processing in reward and salience circuits predicts social communication in children with autism. *ELife*, *8*, e39906.

Ackermann, H., Hage, S. R., & Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behavioral and Brain Sciences*, *37*(6), 529–546.

Adamson, L. B., & Frick, J. E. (2003). The still face: A history of a shared experimental paradigm. *Infancy*, *4*(4), 451–473.

Albert, R. R., Schwade, J. A., & Goldstein, M. H. (2018). The social functions of babbling: Acoustic and contextual characteristics that facilitate maternal responsiveness. *Developmental Science*, *18*(5), e12641–e12611.

Amsel, A. (1958). The role of frustrative nonreward in noncontinuous reward situations. *Psychological Bulletin*, *55*(2), 102–119.

Amsel, A. (1962). Frustrative nonreward in partial reinforcement and discrimination learning: Some recent history and a theoretical extension. *Psychological Review*, *69*(4), 306–328.

Bacon, W. E. (1962). Partial-reinforcement extinction effect following different amounts of training. *Journal of Comparative and Physiological Psychology*, *55*(6), 998–1003.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bigelow, A. E., & Power, M. (2016). Effect of maternal responsiveness on young infants' social bidding-like behavior during the still face task. *Infant and Child Development*, *25*(3), 256–276.

Bigelow, A. E., Power, M., Bulmer, M., & Gerrior, K. (2018). The effect of maternal mirroring behavior on infants' early social bidding during the still-face task. *Infancy*, *23*(3), 367–385.

Bigelow, A. E., & Rochat, P. (2006). Two-month-old infants' sensitivity to social contingency in mother–infant and stranger–infant interaction. *Infancy*, *9*(3), 313–325.

Bornstein, M. H., Arterberry, M. E., & Mash, C. (2004). Long-term memory for an emotional interpersonal interaction occurring at 5 months of age. *Infancy*, *6*(3), 407–416.

Bourvis, N., Singer, M., Georges, C. S., Bodeau, N., Chetouani, M., Cohen, D., & Feldman, R. (2018). Pre-linguistic infants employ complex communicative loops to engage mothers in social exchanges and repair interaction ruptures. *Royal Society Open Science*, *5*(1), 170274.

Corps, R. E., Gambi, C., & Pickering, M. J. (2017). Coordinating utterances during turn-taking: The role of prediction, response preparation, and articulation. *Discourse Processes*, *55*(2), 1–11.

DeCasper, A., & Fifer, W. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*(4448), 1174–1176.

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development*, *9*(2), 133–150.

Delgado, C. E. F., Messinger, D. S., & Yale, M. E. (2002). Infant responses to direction of parental gaze: A comparison of two still-face conditions. *Infant Behavior and Development*, *25*(3), 311–318.

Donnellan, E., Bannard, C., McGillion, M. L., Slocombe, K. E., & Matthews, D. (2020). Infants' intentionally communicative vocalizations elicit responses from caregivers and are the best predictors of the transition to language: A longitudinal investigation of infants' vocalizations, gestures and word production. *Developmental Science*, *23*(1), e12843.

Dorfman, H. M., & Gershman, S. J. (2019). Controllability governs the balance between Pavlovian and instrumental action selection. *Nature Communications*, *10*(1), 5826.

Dunsmoor, J. E., Niv, Y., Daw, N., & Phelps, E. A. (2015). Rethinking extinction. *Neuron*, *88*(1), 47–63.

Ekas, N. V., Haltigan, J. D., & Messinger, D. S. (2013). The dynamic still-face effect: Do infants decrease bidding over time when parents are not responsive? *Developmental Psychology*, *49*(6), 1027–1035.

Elmlinger, S. L., Goldstein, M., & Casillas, M. (2022). Immature vocalizations simplify the speech of Tseltal Mayan and US caregivers. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*, 1256–1262.

Elmlinger, S. L., Park, D., Schwade, J. A., & Goldstein, M. H. (2021). Comparing word diversity versus amount of speech in parents' responses to infants' prelinguistic vocalizations. *IEEE Transactions on Cognitive and Developmental Systems*, *Early Access*, 1.

Elmlinger, S. L., Schwade, J. A., & Goldstein, M. H. (2019a). The ecology of prelinguistic vocal learning: Parents simplify the structure of their speech in response to babbling. *Journal of Child Language*, *46*(5), 998–1011.

Elmlinger, S. L., Schwade, J. A., & Goldstein, M. H. (2019b). Babbling elicits simplified caregiver speech: Findings from natural interaction and simulation. In: 2019 Joint IEEE 9th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob), 1–6.

Fagan, M. K., & Doveikis, K. N. (2017). Ordinary interactions challenge proposals that maternal verbal responses shape infant vocal development. *Journal of Speech, Language, and Hearing Research*, *60*(10), 2819–2827.

Faust, K., Carouso-Peck, S., Elson, M., & Goldstein, M. (2020). The origins of social knowledge in altricial species. *Annual Review of Developmental Psychology*, *2*, 225–246.

Festinger, L. (1961). The psychological effects of insufficient rewards. *American Psychologist*, *16*(1), 1–11.

Field, T. M., Cohen, D., Garcia, R., & Greenberg, R. (1984). Mother-stranger face discrimination by the newborn. *Infant Behavior and Development*, *7*(1), 19–25.

Franklin, B., Warlaumont, A. S., Messinger, D., Bene, E., Iyer, S. N., Lee, C.-C., Lambert, B., & Oller, D. K. (2013). Effects of parental interaction on infant vocalization rate, variability and vocal type. *Language Learning and Development*, *10*(3), 279–296.

Gilkerson, J., Richards, J. A., Warren, S. F., Montgomery, J. K., Greenwood, C. R., Oller, D. K., Hansen, J. H. L., & Paul, T. D. (2017). Mapping the early language environment using all-day recordings and automated analysis. *American Journal of Speech-Language Pathology*, *26*(2), 248–265.

Goldstein, M. H., & Brodsky, P. (2006). *EventCoder [Computer software]*. Ithaca, NY: Authors.

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences*, *100*(13), 8030–8035.

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*(5), 515–523.

Goldstein, M. H., & Schwade, J. A. (2010). From birds to words: Perception of structure in social interactions guides vocal development and language learning. In: (M. S. Blumberg, J. H. Freeman, & S. R. Robinson Eds.), *The Oxford handbook of developmental behavioral neuroscience* (pp. 708–729). Oxford University Press.

Goldstein, M. H., Schwade, J. A., & Bornstein, M. H. (2009). The value of vocalizing: Five-month-old infants associate their own noncry vocalizations with responses from caregivers. *Child Development*, *80*(3), 636–644.

Goldstein, M. H., & West, M. J. (1999). Consistent responses of human mothers to prelinguistic infants: The effect of prelinguistic repertoire size. *Journal of Comparative Psychology*, *113*(1), 52–58.

Gómez, R., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, *7*(2), 183–206.

Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., & Parlato-Oliveira, E. (2015). Early development of turn-taking in vocal interaction between mothers and infants. *Frontiers in Psychology*, *6*, 1167.

Gros-Louis, J., West, M. J., Goldstein, M. H., & King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development*, *30*(6), 509–516.

Guenther, F. H. (2016). *Neural control of speech*. MIT Press.

Ha, S., Johnson, C. J., Oller, K. D., & Yoo, H. (2021). Cross-linguistic comparison of utterance shapes in Korean- and English-learning children: An ambient language effect. *Infant Behavior and Development*, *62*, 101528.

Holton, R. B. (1961). Amplitude of an instrumental response following the cessation of reward. *Child Development*, *32*(1), 107.

Hsu, H. C., & Jeng, S. F. (2008). Two-month-olds' attention and affective response to maternal still face: A comparison between term and preterm infants in Taiwan. *Infant Behavior and Development*, *31*(2), 194–206.

Iyer, S. N., Denson, H., Lazar, N., & Oller, D. K. (2016). Volubility of the human infant: Effects of parental interaction (or lack of it). *Clinical Linguistics & Phonetics*, *30*(6), 1–19.

Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., Jasnow, M. D., Rochat, P., & Stern, D. N. (2001). Rhythms of dialogue in infancy: Coordinated timing in development. *Monographs of the Society for Research in Child Development*, *66*, i–149.

Janik, V. M., & Slater, P. J. B. (2000). The different roles of social learning in vocal communication. *Animal Behaviour*, *60*(1), 1–11.

Jhang, Y., & Oller, D. K. (2017). Emergence of functional flexibility in infant vocalizations of the first 3 months. *Frontiers in Psychology*, *8*, 300.

Jones, S. S., Collins, K., & Hong, H. (1991). An audience effect on smile production in 10-month-old infants. *Psychological Science*, *2*, 45–49.

Kaplan, F., & Oudeyer, P.-Y. (2004). *Embodied Artificial Intelligence, International Seminar*. Germany: Dagstuhl Castle, pp. 259–270. July 7–11, 2003. Revised Papers.

Kidd, C., & Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron*, *88*(3), 449–460.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *Plos One*, *7*(5), e36399.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The Goldilocks effect in infant auditory attention. *Child Development*, *85*(5), 1795–1804.

Kremin, L. V., Alves, J., Orena, A. J., Polka, L., & Byers-Heinlein, K. (2021). Code-switching in parents' everyday speech to bilingual infants. *Journal of Child Language*, *First View*, 1–27.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). *Emmeans: Estimated marginal means, aka least-squares means*. R package version, 1(1), 3.

Levinson, S. C. (2016). Turn-taking in human communication – Origins and implications for language processing. *Trends in Cognitive Sciences*, *20*(1), 6–14.

Ligneul, R. (2021). Prediction or causation? Towards a redefinition of task controllability. *Trends in Cognitive Sciences*, *25*(6), 431–433.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*(4), 276.

Maier, S. F., & Seligman, M. E. P. (2016). Learned helplessness at fifty: Insights from neuroscience. *Psychological Review*, *123*(4), 349–367.

Masek, L. R., McMillan, B. T. M., Paterson, S. J., Tamis-LeMonda, C. S., Golinkoff, R. M., & Hirsh-Pasek, K. (2021). Where language meets attention: How contingent interactions promote learning. *Developmental Review*, *60*, 100961.

McConnell, B. L., & Miller, R. R. (2014). Associative accounts of recovery-from-extinction effects. *Learning and Motivation*, *46*, 1–15.

McLeod, S., & Crowe, K. (2018). Children's consonant acquisition in 27 languages: A cross-linguistic review. *American Journal of Speech-Language Pathology*, *27*(4), 1546–1571.

Mesman, J., van IJzendoorn, M. H., & Bakermans-Kranenburg, M. J. (2009). The many faces of the still-face paradigm: A review and meta-analysis. *Developmental Review*, *29*(2), 120–162.

Messinger, D. (2005). Smiling. In: (N. J. Salkind Ed.), *The encyclopedia of human development* (vol., 3, pp. 1183–1185). Thousand Oaks, CA: Sage.

Messinger, D. M., Ruvolo, P., Ekas, N. V., & Fogel, A. (2010). Applying machine learning to infant interaction: The development is in the details. *Neural Networks*, *23*(8-9), 1004–1016.

Messinger, D. S. (2002). Positive and negative: Infant facial expressions and emotions. *Current Directions in Psychological Science*, *11*(1), 1–6.

Miller, J. L., Ables, E. M., King, A. P., & West, M. J. (2009). Different patterns of contingent stimulation differentially affect attention span in prelinguistic infants. *Infant Behavior and Development*, *32*(3), 254–261.

Miller, J. L., & Gros-Louis, J. (2013). Socially guided attention influences infants' communicative behavior. *Infant Behavior and Development*, *36*(4), 627–634.

Movellan, J. R., & Watson, J. S. (1987). Perception of directional attention. Infant Behavior and Development: Abstracts of the Sixth International Conference of Infant Studies.

Nadel, J., Carchon, I., Kervella, C., Marcelli, D., & Réserbat-Plantey, D. (1999). Expectancies for social contingency in 2-month-olds. *Developmental Science*, *2*(2), 164–173.

Northrup, J. B., Ridley, J., Foley, K., Moses-Kolko, E. L., Keenan, K., & Hipwell, A. E. (2019). A transactional model of infant still-face response and maternal behavior during the first year. *Infancy*, *24*(5), 787–806.

Oller, D. K. (2000). *The emergence of the speech capacity*. Mahwah, NJ: Lawrence Erlbaum and Associates.

Oller, D. K., Buder, E. H., Ramsdell, H. L., Warlaumont, A. S., Chorna, L., & Bakeman, R. (2013). Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Sciences*, *110*(16), 6318–6323.

Oller, D. K., Caskey, M., Yoo, H., Bene, E. R., Jhang, Y., Lee, C.-C., Bowman, D. D., Long, H. L., Buder, E. H., & Vohr, B. (2019). Preterm and full term infant vocalization and the origin of language. *Scientific Reports*, *9*(1), 14734.

Orena, A. J., & Werker, J. F. (2021). Infants' mapping of new faces to new voices. *Child Development*, *92*(5), e1048–e1060.

Oudeyer, P. -. Y., & Smith, L. B. (2016). How evolution may work through curiosity-driven developmental process. *Topics in Cognitive Science*, *8*(2), 492–502.

Papoušek, M., & Papoušek, H. (1989). Forms and functions of vocal matching in interactions between mothers and their precanonical infants. *First Language*, *9*(6), 137–157.

Pickering, & Garrod (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347.

Raab, H. A., Foord, C., Ligneul, R., & Hartley, C. A. (2022). Developmental shifts in computations used to detect environmental controllability. *PLOS Computational Biology*, *18*(6), e1010120.

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2017). The impact of early social interactions on later language development in Spanish–English bilingual infants. *Child Development*, *88*(4), 1216–1234.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: (A. H. Black & W. F. Prokasy Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Ritwika, V. P. S., Pretzer, G. M., Mendoza, S., Shedd, C., Kello, C. T., Gopinathan, A., & Warlaumont, A. S. (2020). Exploratory dynamics of vocal foraging during infant-caregiver communication. *Scientific Reports*, *10*(1), 10469.

Rollins, P. R. (2003). Caregivers' contingent comments to 9-month-old infants: Relationships with later language. *Applied Psycholinguistics*, *24*(2), 221–234.

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*(2), 339–349.

Ruvolo, P., Messinger, D., & Movellan, J. (2015). Infants time their smiles to make their moms smile. *PLoS One*, *10*(9), e0136492.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599.

Sharpe, M. J., Chang, C. Y., Liu, M. A., Batchelor, H. M., Mueller, L. E., Jones, J. L., Niv, Y., & Schoenbaum, G. (2017). Dopamine transients are sufficient and necessary for acquisition of model-based associations. *Nature Neuroscience*, *20*(5), 735–742.

Striano, T., & Bertin, E. (2005). Coordinated affect with mothers and strangers: A longitudinal analysis of joint engagement between 5 and 9 months of age. *Cognition & Emotion*, *19*(5), 781–790.

Sullivan, M. W., & Lewis, M. (2003). Contextual determinants of anger and other negative expressions in young infants. *Developmental Psychology*, *39*(4), 693.

Syal, S., & Finlay, B. L. (2011). Thinking outside the cortex: Social motivation in the evolution and development of language. *Developmental Science*, *14*(2), 417–430.

Tamis-LeMonda, C. S., & Bornstein, M. H. (2002). Maternal responsiveness and early language acquisition. *Advances in Child Development and Behavior*, *29*, 89–127.

Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development*, *72*(3), 748–767.

Tarabulsy, G. M., Tessier, R., & Kappas, A. (1996). Contingency detection and the contingent organization of behavior in interactions: Implications for socioemotional development in infancy. *Psychological Bulletin*, *120*(1), 25–41.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, *24*(4), 629–640.

Tronick, E., Als, H., Adamson, L., Wise, S., & Brazelton, T. B. (1979). The Infant's response to entrapment between contradictory messages in face-to-face interaction. *Journal of the American Academy of Child Psychiatry*, *17*(1), 1–13.

Tronick, E. Z., Ricks, M., & Cohn, J. F. (1982). Maternal and infant affective exchange: Patterns of adaptation. In: (T. Field & A. Fogel Eds.), *Emotion and early interaction* (pp. 83–100). Hillsdale, NJ: Lawrence Erlbaum.

Vihman, M. M. (2017). Learning words and learning sounds: Advances in language development. *British Journal of Psychology*, *108*(1), 1–27.

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, *10*(2), 159–164.

Vukatana, E., Graham, S. A., Curtin, S., & Zepeda, M. S. (2015). One is not enough: Multiple exemplars facilitate infants' generalizations of novel properties. *Infancy*, *20*(5), 548–575.

Wade, S., & Kidd, C. (2019). The role of prior knowledge and curiosity in learning. *Psychonomic Bulletin & Review*, *31*(2), 1–11.

Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2014). A social feedback loop for speech development and its reduction in autism. *Psychological Science*, *25*(7), 1314–1324.

Watson, J. S. (1967). Memory and "contingency analysis" in infant learning. *Merril-Palmer Quarterly of Behavior and Development*, *13*(1), 55–76.

Wentworth, N., & Haith, M. M. (1992). Event-specific expectations of 2- and 3-month-old infants. *Developmental Psychology*, *28*(5), 842–850.

Yirmiya, N., Gamliel, I., Pilowsky, T., Feldman, R., Baron-Cohen, S., & Sigman, M. (2006). The development of siblings of children with autism at 4 and 14 months: Social engagement, communication, and cognition. *Journal of Child Psychology and Psychiatry*, *47*(5), 511–523.

How to cite this article: Elmlinger, S. L., Schwade, J. A., Vollmer, L., & Goldstein, M. H. (2022). Learning how to learn from social feedback: The origins of early vocal development. *Developmental Science*, e13296. https://doi.org/10.1111/desc.13296